Nathan Wemmer

Code

This is an R Markdown (http://rmarkdown.rstudio.com) Notebook. When you execute code within the notebook, the results appear beneath the code.

Try executing this chunk by clicking the *Run* button within the chunk or by placing your cursor inside it and pressing *Ctrl+Shift+Enter*.

Add a new chunk by clicking the *Insert Chunk* button on the toolbar or by pressing *Ctrl+Alt+I*.

When you save the notebook, an HTML file containing the code and output will be saved alongside it (click the *Preview* button or press *Ctrl+Shift+K* to preview the HTML file).

The preview shows you a rendered HTML copy of the contents of the editor. Consequently, unlike *Knit*, *Preview* does not run any R code chunks. Instead, the output of the chunk when it was last run in the editor is displayed. # Assignment 8: Sections 10.1-10.13 # Section 10.1

Hide

```
setwd("C:/Users/Nathan/Desktop/school/statistical data management/therbook")
reg.data <- read.table("regression.txt",header=T)
```

```
the condition has length > 1 and only the first element will be usedthe condition has length > 1
and only the first element will be used
```

Hide

```
attach(reg.data)
```

```
The following objects are masked from data (pos = 6):

    growth, tannin

The following objects are masked from data (pos = 8):

    growth, tannin
```

Hide

```
names(reg.data)
```

```
[1] "growth" "tannin"
```

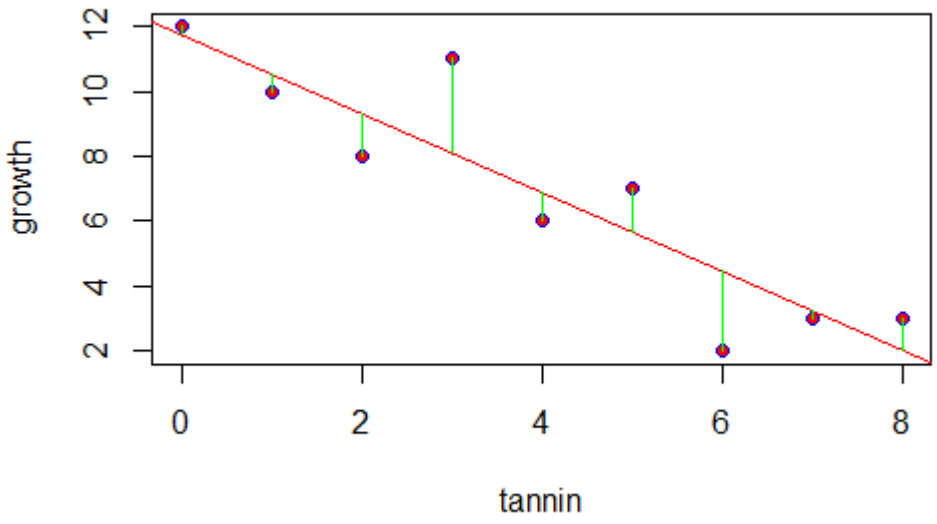
Hide

```
plot(tannin,growth,pch=21,col="blue",bg="red")

model <- lm(growth~tannin)
abline(model,col="red")
```

Hide

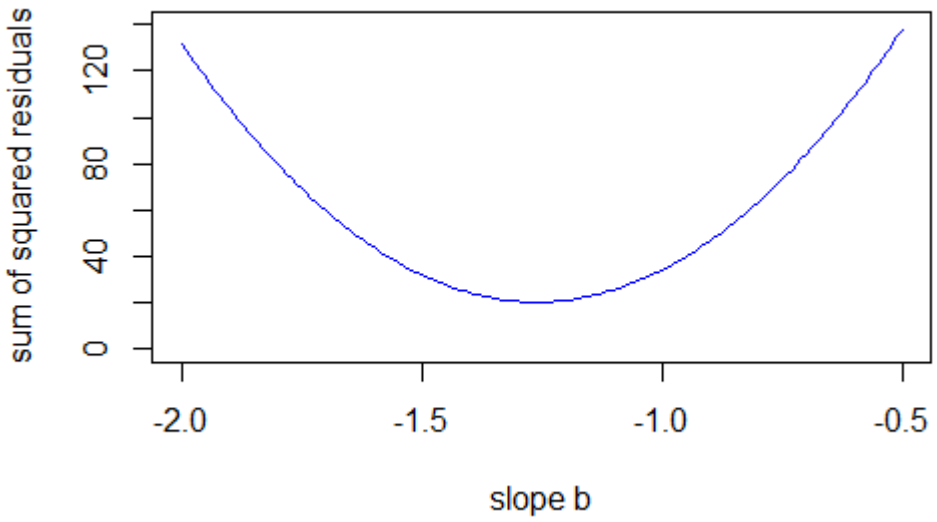
```
yhat <- predict(model,tannin=tannin)
join <- function(i)
  lines(c(tannin[i],tannin[i]),c(growth[i],yhat[i]),col="green")
sapply(1:9,join)
```

```
[[1]]
NULL

[[2]]
NULL

[[3]]
NULL

[[4]]
NULL

[[5]]
NULL

[[6]]
NULL

[[7]]
NULL

[[8]]
NULL

[[9]]
NULL
```



Hide

```
bs <- seq(-2,-0.5,0.01)
SSE <- function(i) sum((growth - 12 - bs[i]*tannin)^2)
plot(bs,sapply(1:length(bs),SSE),type="l",ylim=c(0,140),
     xlab="slope b",ylab="sum of squared residuals",col="blue")
```



Section 10.1.1

Hide

```
sum(tannin);sum(tannin^2);sum(growth);sum(growth^2);sum(tannin*growth)
```

```
[1] 36
[1] 204
[1] 62
[1] 536
[1] 175
```

Hide

```
XY <- cbind(1,growth,tannin)
t(XY) %*% XY
```

```
         growth tannin
        9     62     36
growth 62    536    175
tannin 36    175    204
```

Section 10.1.2

Hide

```
mean(growth)+1.216667*mean(tannin)
```

```
[1] 11.75556
```

Hide

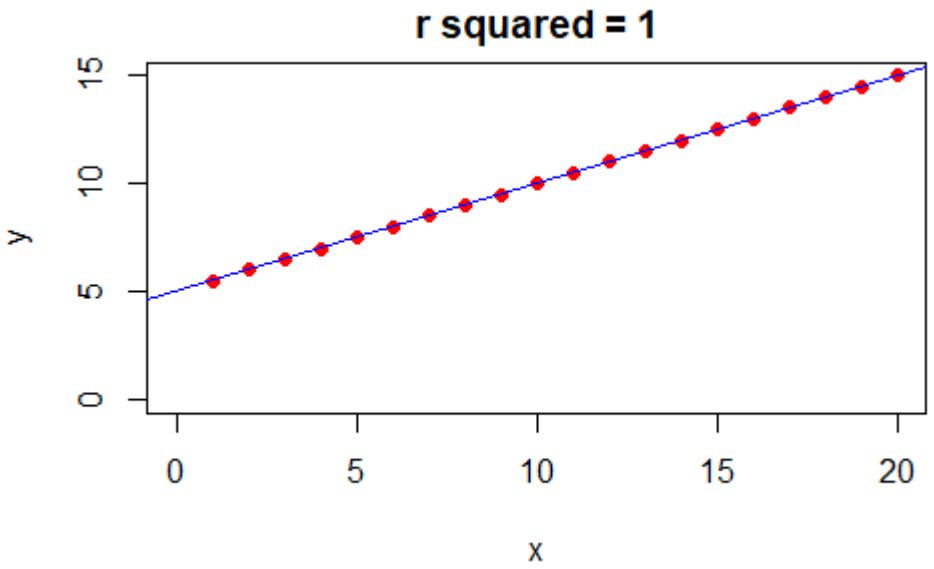
```
lm(growth~tannin)
```

```

Call:
lm(formula = growth ~ tannin)

Coefficients:
(Intercept)       tannin
     11.756       -1.217
```

Section 10.1.3

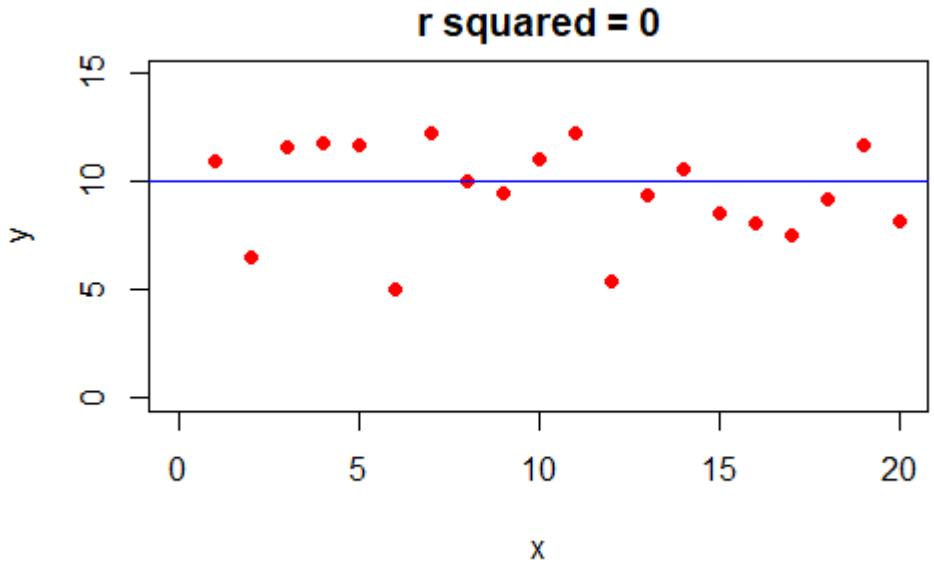
Hide

```
x=1:20
y <- 5+0.5*x
plot(x,y,pch=16,xlim=c(0,20),ylim=c(0,15),col="red",main="r squared = 1")
abline(5,0.5,col="blue")
```



Hide

```
x=1:20
#y <- 5+runif(30)*10
y <- 5+runif(20)*10
plot(x,y,pch=16,xlim=c(0,20),ylim=c(0,15),col="red",main="r squared = 0")
abline(h=10,col="blue")
```



Hide

```
deviance(lm(growth~1))
```

```
[1] 108.8889
```

Hide

```
deviance(lm(growth~tannin))
```

```
[1] 20.07222
```

Hide

```
summary(lm(growth~tannin))[[8]]
```

```
[1] 0.8156633
```

Section 10.1.4

Hide

```
(sse <- deviance(lm(growth~tannin)))
```

```
[1] 20.07222
```

Hide

```
(ssy <- deviance(lm(growth~1)))
```

```
[1] 108.8889
```

Hide

```
(ssr <- ssy-sse)
```

```
[1] 88.81667
```

Hide

```
anova(lm(growth~tannin))
```

```
Analysis of Variance Table

Response: growth
          Df Sum Sq Mean Sq F value    Pr(>F)
tannin     1 88.817  88.817  30.974 0.0008461 ***
Residuals  7 20.072   2.867
---
Signif. codes:
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Hide

```
qf(0.95,1,7)
```

```
[1] 5.591448
```

Hide

```
1-pf(30.974,1,7)
```

```
[1] 0.0008460725
```

Section 10.1.5

Hide

```
summary(lm(growth~tannin))
```

```
Call:
lm(formula = growth ~ tannin)

Residuals:
    Min      1Q  Median      3Q     Max
-2.4556 -0.8889 -0.2389  0.9778  2.8944

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  11.7556     1.0408  11.295 9.54e-06 ***
tannin       -1.2167     0.2186  -5.565 0.000846 ***
---
Signif. codes:
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.693 on 7 degrees of freedom
Multiple R-squared:  0.8157,    Adjusted R-squared:  0.7893
F-statistic: 30.97 on 1 and 7 DF,  p-value: 0.0008461
```

Hide

```
confint(model)
```

```
               2.5 %     97.5 %
(Intercept)  9.294457 14.2166544
tannin      -1.733601 -0.6997325
```

Section 10.1.6

Hide

```
model <- lm(growth~tannin)
predict(model,list(tannin=5.5))
```

```
       1
5.063889
```

Hide

```
predict(model,list(tannin=c(3.3,4.4,5.5,6.6)))
```

```
       1        2        3        4
7.740556 6.402222 5.063889 3.725556
```

Section 10.1.7

Hide

```
windows(7,7)
par(mfrow=c(2,2))
plot(model)
model2 <- update(model,subset=(tannin != 6))
summary(model2)
```

```

Call:
lm(formula = growth ~ tannin, subset = (tannin != 6))

Residuals:
    Min      1Q  Median      3Q     Max 
-1.4549 -0.9572 -0.1622  0.4572  2.6622 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  11.6892     0.8963  13.042 1.25e-05 ***
tannin       -1.1171     0.1956  -5.712  0.00125 ** 
---
Signif. codes:  
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.457 on 6 degrees of freedom
Multiple R-squared:  0.8446,    Adjusted R-squared:  0.8188 
F-statistic: 32.62 on 1 and 6 DF,  p-value: 0.001247
```

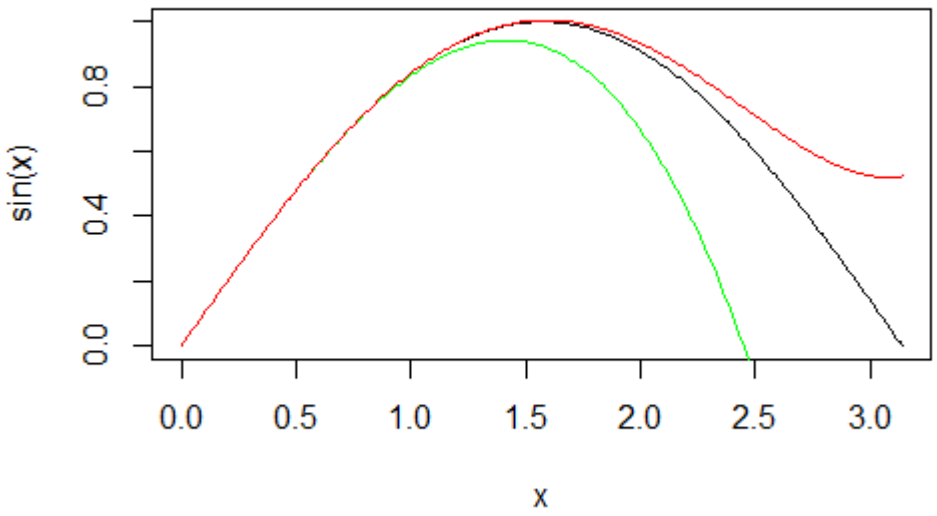
Section 10.2

Hide

```
x <- seq(0,pi,0.01)
y <- sin(x)
plot(x,y,type="l",ylab="sin(x)")

a1 <- x-x^3/factorial(3)
lines(x,a1,col="green")
```

Hide

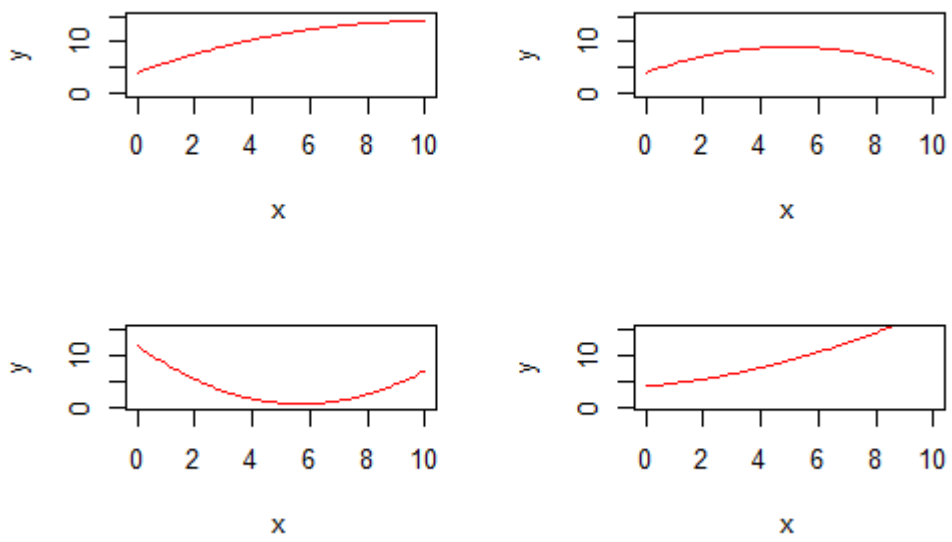
```
a2 <- x-x^3/factorial(3)+x^5/factorial(5)
lines(x,a2,col="red")
```



I had a problem here # Section 10.3

Hide

```
par(mfrow=c(2,2))
x <- seq(0,10,0.1)
y1 <- 4 + 2 * x - 0.1 * x^2
y2 <- 4 + 2 * x - 0.2 * x^2
y3 <- 12 - 4 * x + 0.35 * x^2
y4 <- 4 + 0.5 * x + 0.1 * x^2
plot(x,y1,type="l",ylim=c(0,15),ylab="y",col="red")
```



Hide

```
plot(x,y2,type="l",ylim=c(0,15),ylab="y",col="red")
plot(x,y3,type="l",ylim=c(0,15),ylab="y",col="red")
```

Hide

```
plot(x,y4,type="l",ylim=c(0,15),ylab="y",col="red")

setwd("C:/Users/Nathan/Desktop/school/statistical data management/therbook")
```

```
The working directory was changed to C:/Users/Nathan/Desktop/school/statistical data management/therbook inside a notebook chunk. The working directory will be reset when the chunk is finished running. Use the knitr root.dir option in the setup chunk to change the working directory for notebook chunks.
```

Hide

```
poly <- read.table("diminish.txt",header=T)
attach(poly)
```

```
The following objects are masked _by_ .GlobalEnv:

    xv, yv

The following objects are masked from poly (pos = 3):

    xv, yv

The following objects are masked from poly (pos = 4):

    xv, yv

The following objects are masked from poly (pos = 5):

    xv, yv

The following objects are masked from poly (pos = 6):

    xv, yv

The following objects are masked from poly (pos = 7):

    xv, yv

The following objects are masked from poly (pos = 8):

    xv, yv
```

Hide

```
names(poly)
```

```
[1] "xv" "yv"
```

Hide

```
windows(7,4)
```

Hide

```
par(mfrow=c(1,2))
model1 <- lm(yv~xv)
plot(xv,yv,pch=21,col="brown",bg="yellow")
```

```
Error in xy.coords(x, y, xlabel, ylabel, log) :
  'x' and 'y' lengths differ
```

Section 10.4

Hide

```
setwd("C:/Users/Nathan/Desktop/school/statistical data management/therbook")
```

```
The working directory was changed to C:/Users/Nathan/Desktop/school/statistical data management/therbook inside a notebook chunk. The working directory will be reset when the chunk is finished running. Use the knitr root.dir option in the setup chunk to change the working directory for notebook chunks.
```

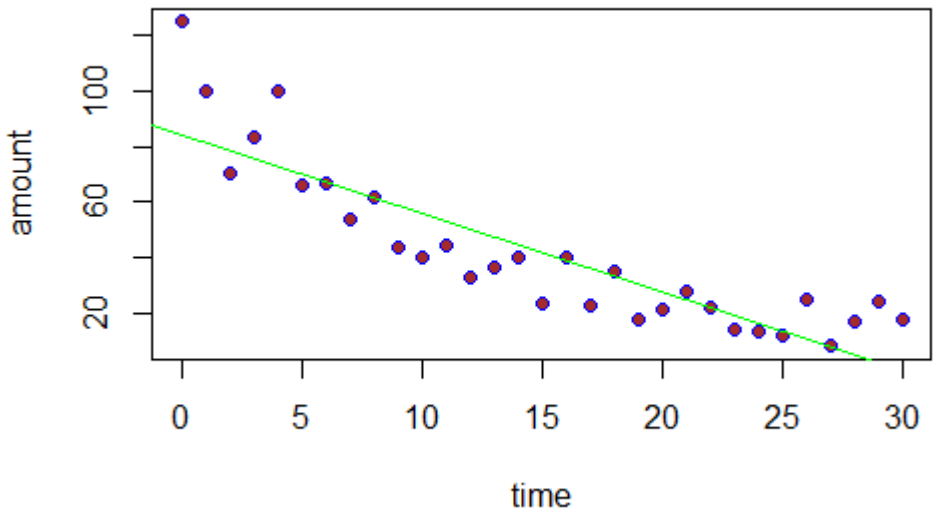
Hide

```
data <- read.table("Decay.txt",header=T)
names(data)
```

```
[1] "time"   "amount"
```

Hide

```
attach(data)
plot(time,amount,pch=21,col="blue",bg="brown")
abline(lm(amount~time),col="green")
```



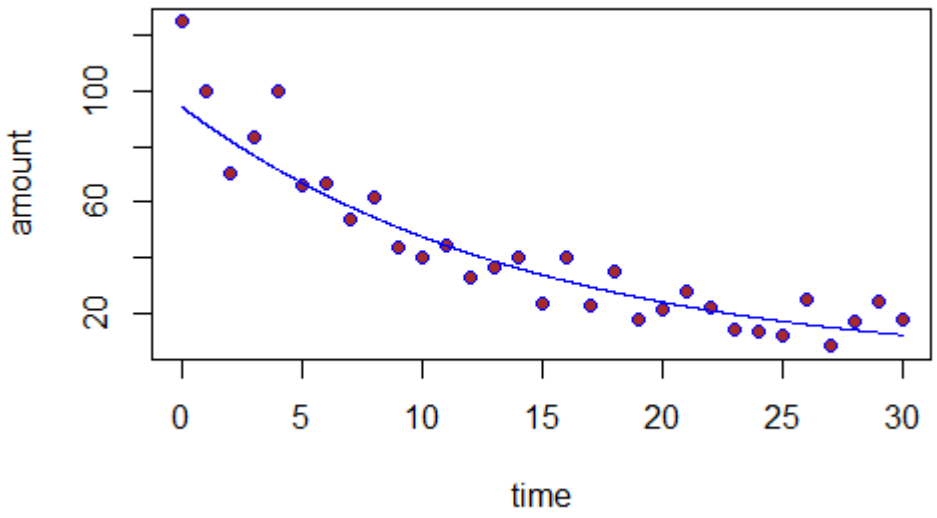
Hide

```
model <- lm(log(amount)~time)
summary(model)
```

```
Call:
lm(formula = log(amount) ~ time)

Residuals:
    Min      1Q  Median      3Q     Max
-0.5935 -0.2043  0.0067  0.2198  0.6297

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  4.547386   0.100295   45.34  < 2e-16 ***
time        -0.068528   0.005743  -11.93 1.04e-12 ***
---
Signif. codes:
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.286 on 29 degrees of freedom
Multiple R-squared:  0.8308,    Adjusted R-squared:  0.825
F-statistic: 142.4 on 1 and 29 DF,  p-value: 1.038e-12
```

Hide

```
ts <- seq(0,30,0.02)
left <- exp(predict(model,list(time=ts)))
plot(time,amount,pch=21,col="blue",bg="brown")
lines(ts,left,col="blue")
```



Section 10.5

Hide

```
setwd("C:/Users/Nathan/Desktop/school/statistical data management/therbook")
power <- read.table("power.txt",header=T)
```

```
the condition has length > 1 and only the first element will be usedthe condition has length > 1
and only the first element will be used
```

Hide

```
attach(power)
```

```
The following object is masked from data (pos = 6):

    response
```

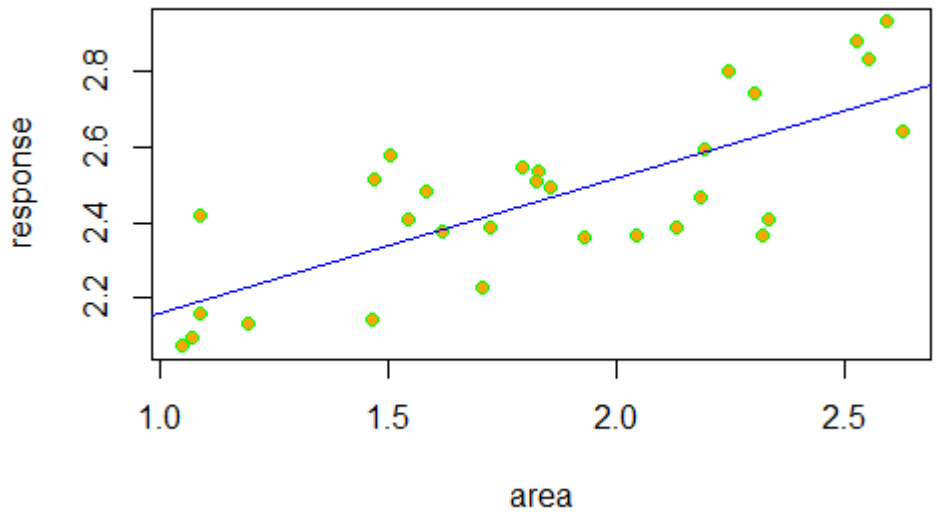
Hide

```
names(power)
```

```
[1] "area"     "response"
```

Hide

```
plot(area,response,pch=21,col="green",bg="orange")
abline(lm(response~area),col="blue")
```



Hide

```
plot(log(area),log(response),pch=21,col="green",bg="orange")
abline(lm(log(response)~log(area)),col="blue")
```

Hide

```
model1 <- lm(response~area)
model2 <- lm(log(response)~log(area))
summary(model2)
```

```
Call:
lm(formula = log(response) ~ log(area))

Residuals:
      Min        1Q    Median        3Q       Max 
-0.100937 -0.043289 -0.000562  0.046095  0.108453 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  0.75378    0.02613  28.843  < 2e-16 ***
log(area)    0.24818    0.04083   6.079 1.48e-06 ***
---
Signif. codes:  
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.06171 on 28 degrees of freedom
Multiple R-squared:  0.5689,    Adjusted R-squared:  0.5535 
F-statistic: 36.96 on 1 and 28 DF,  p-value: 1.48e-06
```

Hide

```
windows(7,7)
plot(area,response,pch=21,col="green",bg="orange")
```

Hide

```
abline(lm(response~area),col="blue")
xv <- seq(1,2.7,0.01)
yv <- exp(0.75378)*xv^0.24818
lines(xv,yv,col="red")
```

Hide

```
plot(area,response,xlim=c(0,5),ylim=c(0,4),pch=21,col="green",bg="orange")
abline(lm(response~area),col="blue")
```

Hide

```
xv <- seq(0,5,0.01)
yv <- exp(0.75378)*xv^0.24818
lines(xv,yv,col="red")
```

Section 10.6

Hide

```
setwd("C:/Users/Nathan/Desktop/school/statistical data management/therbook")
```

```
The working directory was changed to C:/Users/Nathan/Desktop/school/statistical data management/therbook inside a notebook chunk. The working directory will be reset when the chunk is finished running. Use the knitr root.dir option in the setup chunk to change the working directory for notebook chunks.
```

Hide

```
reg.data <- read.table("regression.txt",header=T)
attach(reg.data)
names(reg.data)
```

```
[1] "growth" "tannin"
```

Hide

```
plot(tannin,growth,pch=21,col="blue",bg="red")

model <- lm(growth~tannin)
abline(model,col="blue")
```

Hide

```
coef(model)[2]
```

```
   tannin
-1.216667
```

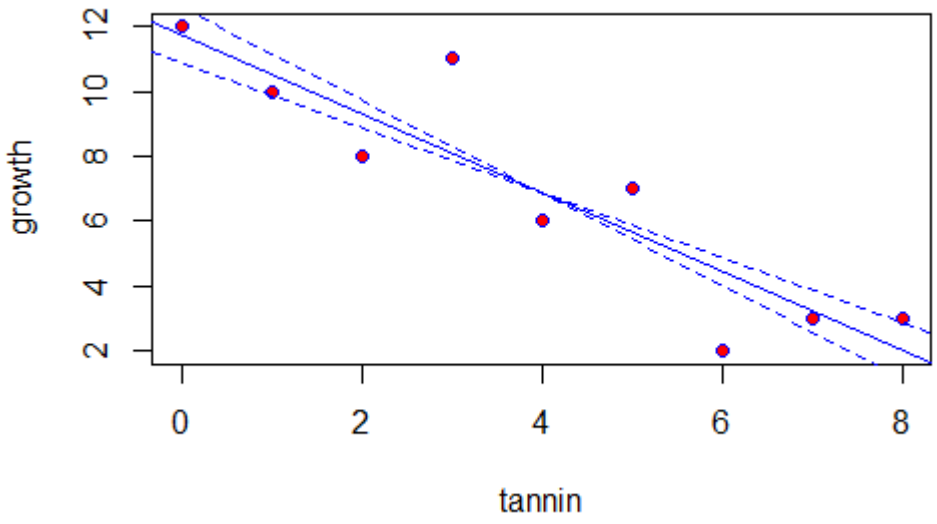
Hide

```
summary(model)[[4]][4]
```

```
[1] 0.2186115
```

Hide

```
se.lines <- function(model){
  b1 <- coef(model)[2]+ summary(model)[[4]][4]
  b2 <- coef(model)[2]- summary(model)[[4]][4]
  xm <- sapply(model[[12]][2],mean)
  ym <- sapply(model[[12]][1],mean)
  a1 <- ym-b1*xm
  a2 <- ym-b2*xm
  abline(a1,b1,lty=2,col="blue")
  abline(a2,b2,lty=2,col="blue")
}
se.lines(model)
```

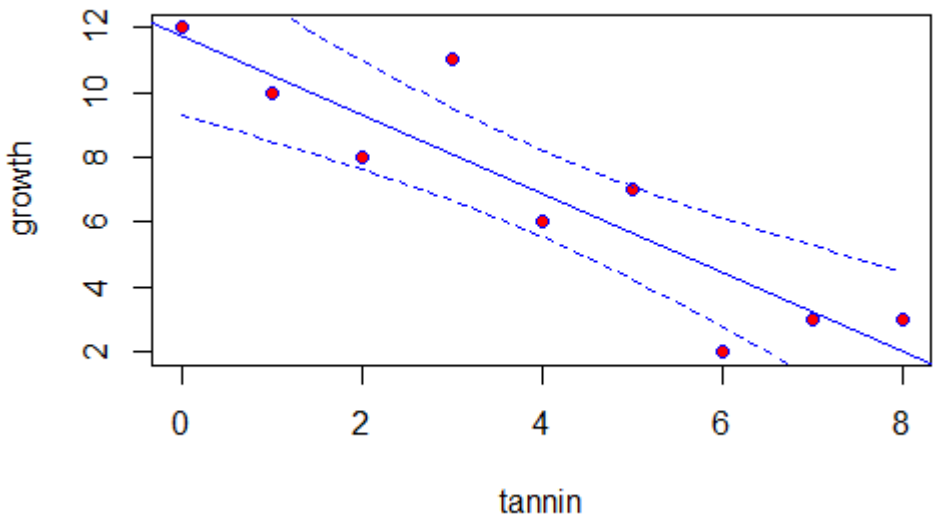


Hide

```
ci.lines <- function(model){
  xm <- sapply(model[[12]][2],mean)
  n <- sapply(model[[12]][2],length)
  ssx <- sum(model[[12]][2]^2)-sum(model[[12]][2])^2/n
  s.t <- qt(0.975,(n-2))
  xv <- seq(min(model[[12]][2]),max(model[[12]][2]),length=100)
  yv <- coef(model)[1]+coef(model)[2]*xv
  se <- sqrt(summary(model)[[6]]^2*(1/n+(xv-xm)^2/ssx))
  ci <- s.t*se
  uyv <- yv+ci
  lyv <- yv-ci
  lines(xv,uyv,lty=2,col="blue")
  lines(xv,lyv,lty=2,col="blue")
}

plot(tannin,growth,pch=21,col="blue",bg="red")
abline(model, col="blue")
```

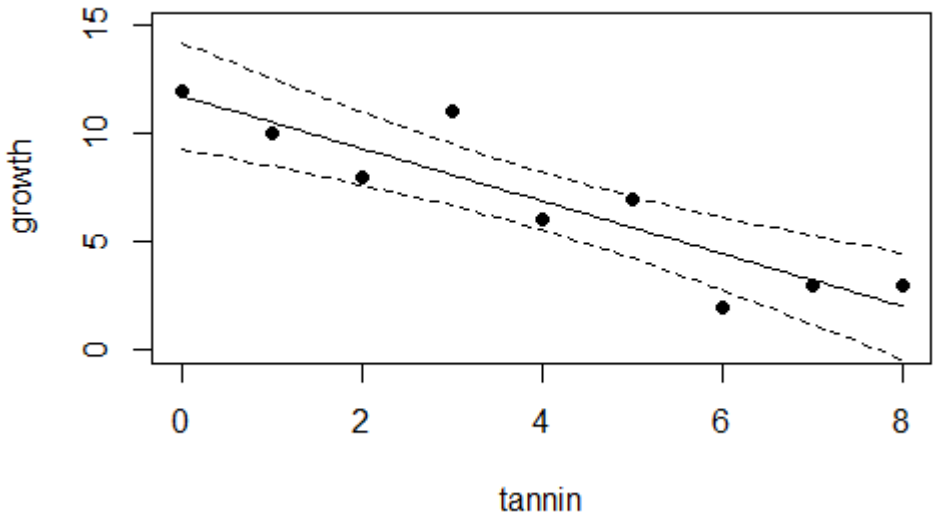
Hide

```
ci.lines(model)
```



Hide

```
plot(tannin,growth,pch=16,ylim=c(0,15))
model <-lm(growth~tannin)

xv <- seq(0,8,0.1)
yv <- predict(model,list(tannin=xv),int="c")
matlines(xv,yv,lty=c(1,2,2),col="black")
```



Section 10.7

Hide

```
setwd("C:/Users/Nathan/Desktop/school/statistical data management/therbook")
data <- read.delim("lackoffit.txt")


attach(data)
names(data)
```

```
[1] "conc" "rate"
```

Hide

```
plot(conc,jitter(rate),pch=16,col="red",ylim=c(0,8),ylab="rate")
abline(lm(rate~conc),col="blue")
```

Hide

```


model.reg <- lm(rate~conc)
summary(model.reg)
```

```
Call:
lm(formula = rate ~ conc)

Residuals:
     Min       1Q   Median       3Q      Max
-1.96429 -0.90476  0.09524  0.27381  2.15476

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   6.7262     0.4559  14.755 7.35e-12 ***
conc         -0.9405     0.1264  -7.439 4.85e-07 ***
---
Signif. codes:
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.159 on 19 degrees of freedom
Multiple R-squared:  0.7444,    Adjusted R-squared:  0.7309
F-statistic: 55.33 on 1 and 19 DF,  p-value: 4.853e-07
```

Hide

```
fac.conc <- factor(conc)
model.aov <- aov(rate~fac.conc)
summary(model.aov)
```

```
            Df Sum Sq Mean Sq F value   Pr(>F)
fac.conc     6  87.81  14.635   17.07 1.05e-05 ***
Residuals   14  12.00   0.857
---
Signif. codes:
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Hide

```
anova(model.reg,model.aov)
```

```
Analysis of Variance Table

Model 1: rate ~ conc
Model 2: rate ~ fac.conc
  Res.Df    RSS Df Sum of Sq      F  Pr(>F)
1     19 25.512
2     14 12.000  5    13.512 3.1528 0.04106 *
---
Signif. codes:
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

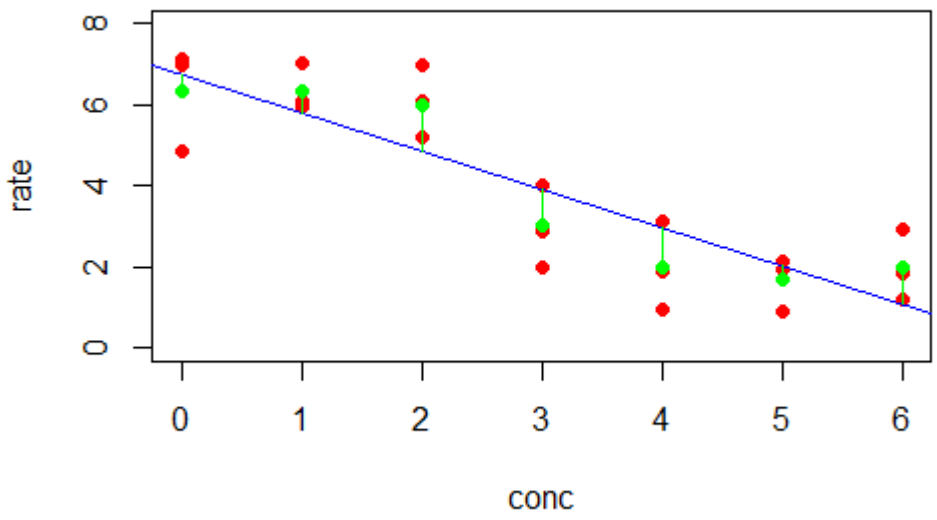
Hide

```
anova(lm(rate~conc+fac.conc))
```

```
Analysis of Variance Table

Response: rate
          Df Sum Sq Mean Sq F value    Pr(>F)
conc       1 74.298  74.298 86.6806 2.247e-07 ***
fac.conc   5 13.512   2.702  3.1528   0.04106 *
Residuals 14 12.000   0.857
---
Signif. codes:
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Hide

```
my <- as.vector(tapply(rate,fac.conc,mean))
for (i in 0:6)
  lines(c(i,i),c(my[i+1],predict(model.reg,list(conc=0:6))[i+1]),col="green")
points(0:6,my,pch=16,col="green")
```



Hide

```
NA
NA
NA
```

Section 10.8

Hide

```
setwd("C:/Users/Nathan/Desktop/school/statistical data management/therbook")
```

```
The working directory was changed to C:/Users/Nathan/Desktop/school/statistical data management/therbook inside a notebook chunk. The working directory will be reset when the chunk is finished running. Use the knitr root.dir option in the setup chunk to change the working directory for notebook chunks.
```

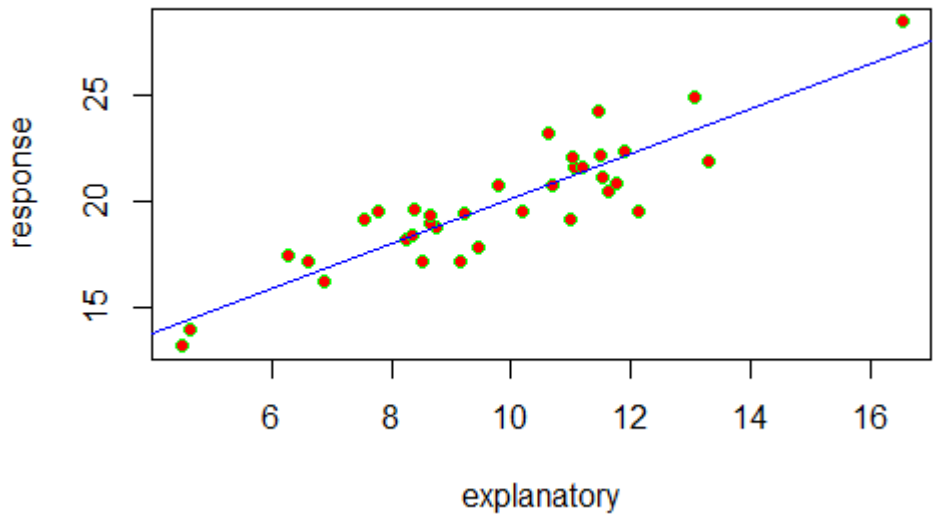
Hide

```
regdat <- read.table("regdat.txt",header=T)

attach(regdat)
names(regdat)
```

```
[1] "explanatory" "response"
```

Hide

```
plot(explanatory,response,pch=21,col="green",bg="red")
model <- lm(response~explanatory)
abline(model,col="blue")
```



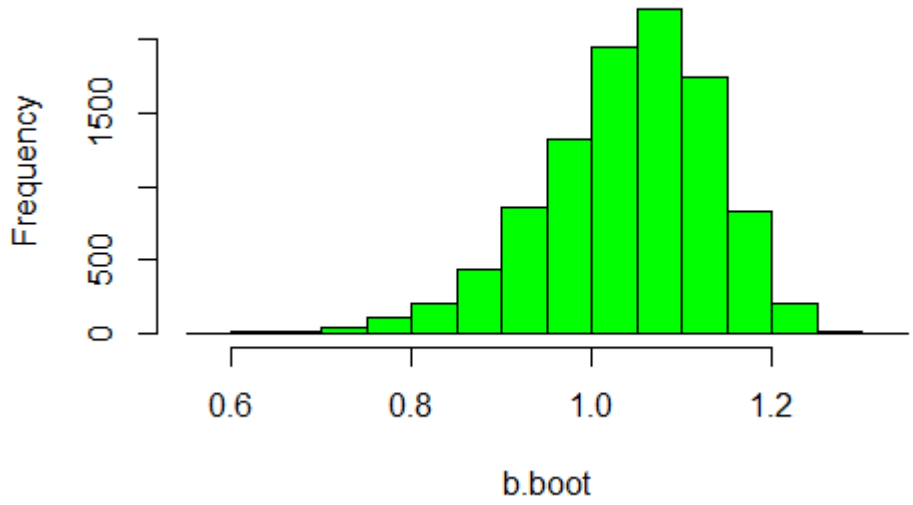
Hide

```
model
```

```

Call:
lm(formula = response ~ explanatory)

Coefficients:
(Intercept)  explanatory
      9.630        1.051
```

Hide

```
b.boot <- numeric(10000)
for (i in 1:10000){
  indices <- sample(1:35,replace=T)
  xv <- explanatory[indices]
  yv <- response[indices]
  model <- lm(yv~xv)
  b.boot[i] <- coef(model)[2]
}
hist(b.boot,main="",col="green")
```



Hide

```
quantile(b.boot,c(0.025,0.975))
```

```
     2.5%     97.5% 
0.8162186 1.1969887 
```

Hide

```
library(boot)
reg.boot <- function(regdat, index){
  xv <- explanatory[index]
  yv <- response[index]
  model <- lm(yv~xv)
  coef(model)
}

reg.model <- boot(regdat,reg.boot,R=10000)
boot.ci(reg.model,index=2)
```

```
bootstrap variances needed for studentized intervals
```

```
BOOTSTRAP CONFIDENCE INTERVAL CALCULATIONS
Based on 10000 bootstrap replicates

CALL :
boot.ci(boot.out = reg.model, index = 2)

Intervals :
Level      Normal              Basic
95%   ( 0.872,  1.254 )   ( 0.904,  1.284 )

Level     Percentile            BCa
95%   ( 0.818,  1.198 )   ( 0.829,  1.203 )
Calculations and Intervals on Original Scale
```

Hide

```
reg.model <- boot(regdat,reg.boot,R=10000)
boot.ci(reg.model,index=2)
```

```
bootstrap variances needed for studentized intervals
```

```
BOOTSTRAP CONFIDENCE INTERVAL CALCULATIONS
Based on 10000 bootstrap replicates

CALL :
boot.ci(boot.out = reg.model, index = 2)

Intervals :
Level      Normal              Basic
95%   ( 0.873,  1.254 )   ( 0.907,  1.285 )

Level     Percentile            BCa
95%   ( 0.816,  1.194 )   ( 0.826,  1.199 )
Calculations and Intervals on Original Scale
```

Hide

```
model <- lm(response~explanatory)
fit <- fitted(model)
res <- resid(model)

residual.boot <- function(res, index){
  y <- fit+res[index]
  model <- lm(y~explanatory)
  coef(model) }

res.model <- boot(res,residual.boot,R=10000)
boot.ci(res.model,index=2)
```

```
bootstrap variances needed for studentized intervals
```

```
BOOTSTRAP CONFIDENCE INTERVAL CALCULATIONS
Based on 10000 bootstrap replicates

CALL :
boot.ci(boot.out = res.model, index = 2)

Intervals :
Level      Normal              Basic
95%   ( 0.875,  1.227 )   ( 0.873,  1.226 )

Level     Percentile            BCa
95%   ( 0.875,  1.229 )   ( 0.870,  1.222 )
Calculations and Intervals on Original Scale
```

Section 10.9

Hide

```
names(regdat)
```

```
[1] "explanatory" "response"
```

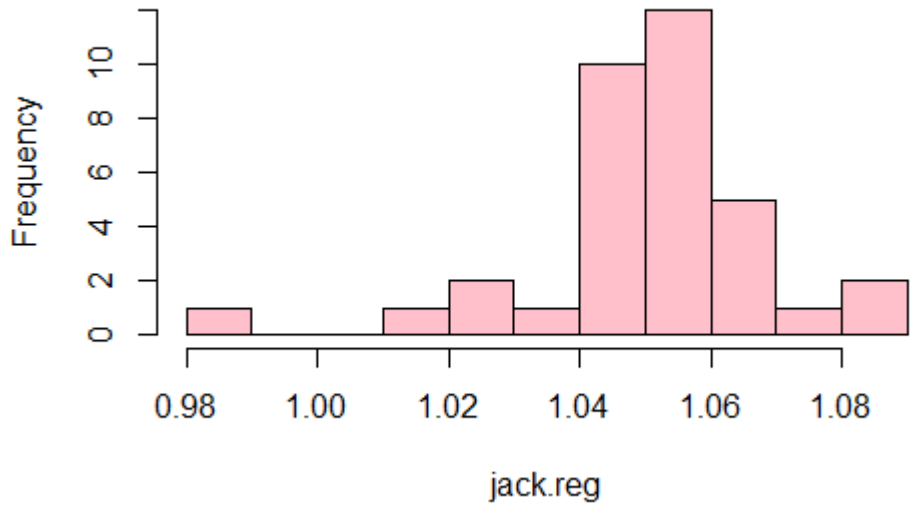
Hide

```
length(response)
```

```
[1] 35
```

Hide

```
jack.reg <- numeric(35)
for (i in 1:35) {
  model <- lm(response[-i]~explanatory[-i])
  jack.reg[i] <- coef(model)[2] }

hist(jack.reg,main="",col="pink")
```



Hide

```
model <- lm(response~explanatory)
which(influence.measures(model)$infmat[,5]
      == max(influence.measures(model)$infmat[,5]))
```

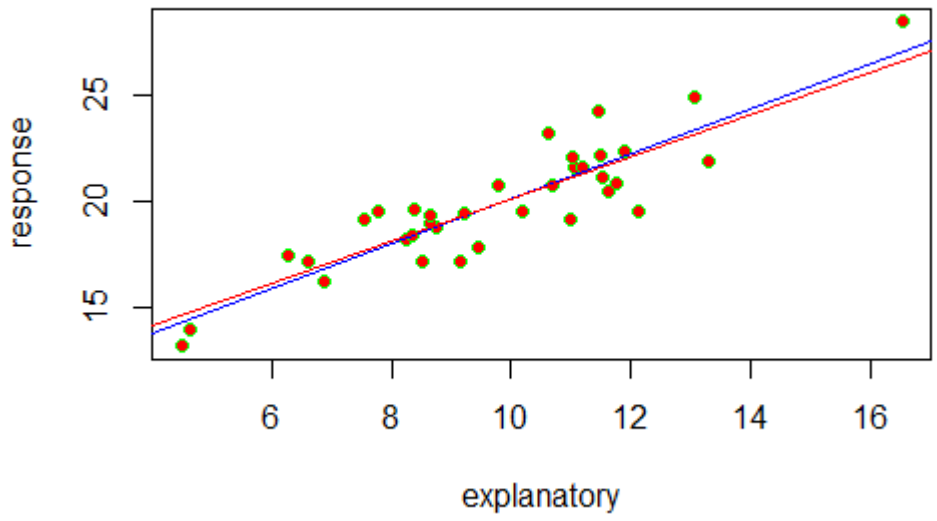
```
22
22
```

Hide

```
plot(explanatory,response,pch=21,col="green",bg="red")
abline(model,col="blue")
```

Hide

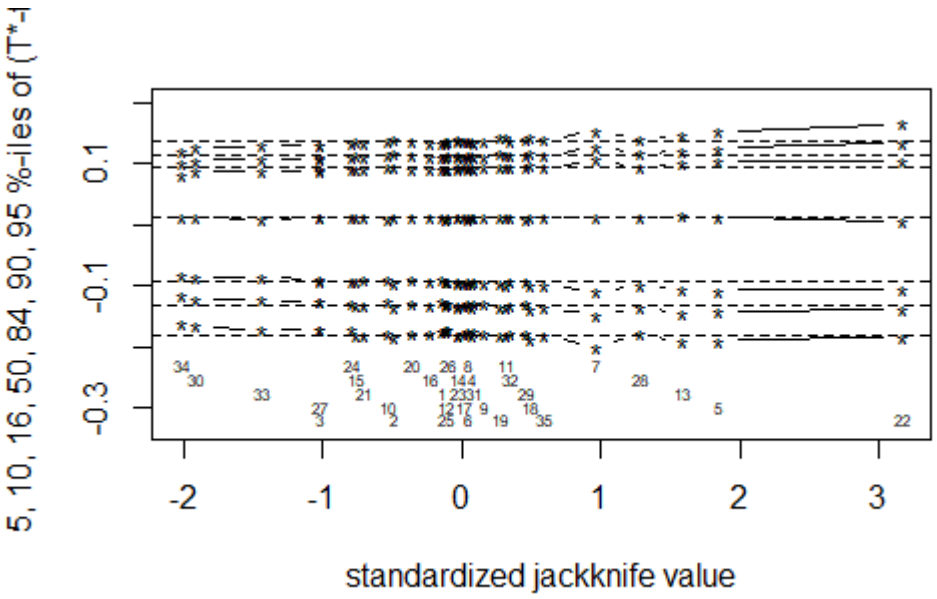
```
abline(lm(response[-22]~explanatory[-22]),col="red")
```



Section 10.10

Hide

```
jack.after.boot(reg.model,index=2)
```



Hide

```
NA
NA
NA
```

Section 10.11

Hide

```
library("car")
```

```
Loading required package: carData
Registered S3 method overwritten by 'data.table':
  method           from
  print.data.table

Attaching package: 勘牺car勘牲

The following object is masked from 勘牺package:boot勘牲:

    logit
```

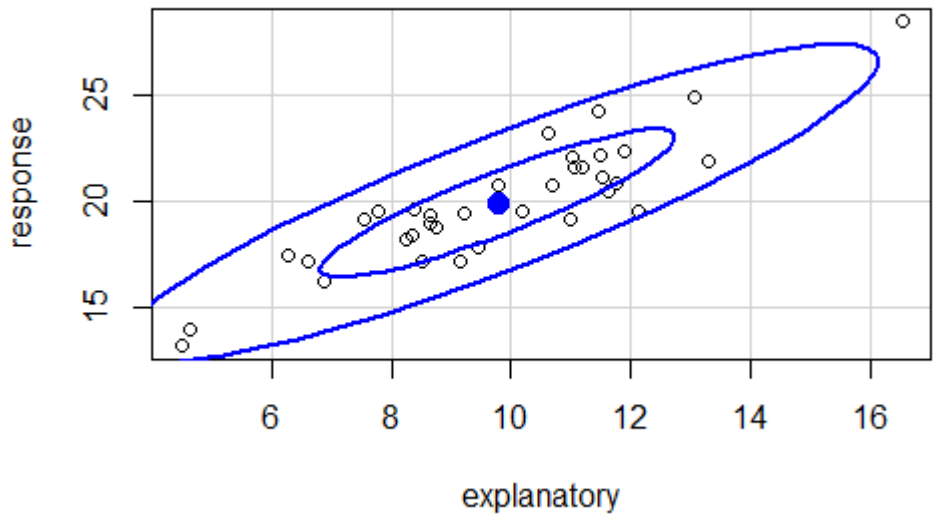
Hide

```
durbinWatsonTest(model)
```

```
 lag Autocorrelation D-W Statistic p-value
   1       0.6066791     0.7846307       0
 Alternative hypothesis: rho != 0
```

Hide

```
dataEllipse(explanatory,response)
```



Section 10.12

Hide

```
setwd("C:/Users/Nathan/Desktop/school/statistical data management/therbook")
```

```
The working directory was changed to C:/Users/Nathan/Desktop/school/statistical data management/
therbook inside a notebook chunk. The working directory will be reset when the chunk is finished
running. Use the knitr root.dir option in the setup chunk to change the working directory for no
tebook chunks.
```

Hide

```
data <- read.table("sasilwood.txt",header=T)
```

```
the condition has length > 1 and only the first element will be usedthe condition has length > 1
and only the first element will be used
```

Hide

```
attach(data)
```

```
The following objects are masked from data (pos = 9):

    Area, Species
```

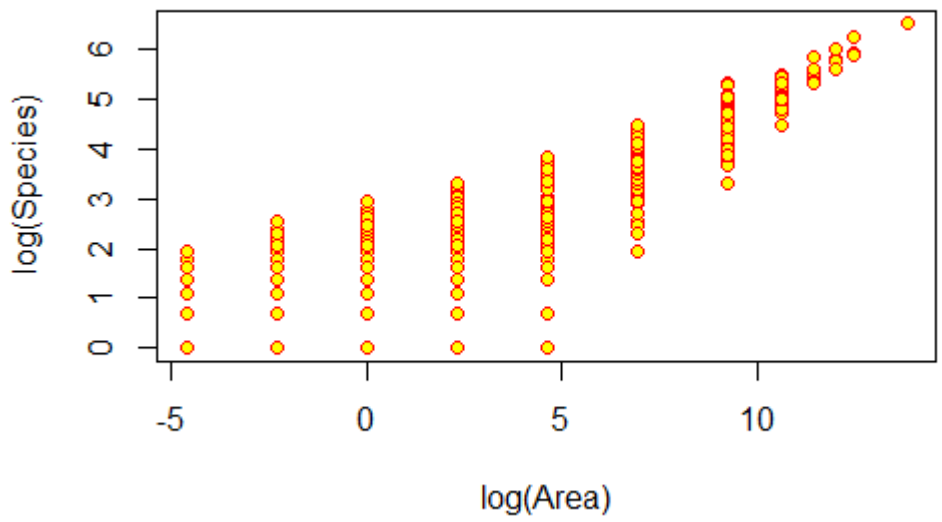
Hide

```
names(data)
```

```
[1] "Species" "Area"
```

Hide

```
plot(log(Species)~log(Area),pch=21,col="red",bg="yellow")
model1 <- lm(log(Species)~log(Area))
par(mfrow=c(2,2))
```



Hide

```
plot(model1)

table(Area)
```

```
Area
  0.01    0.1      1     10    100   1000  10000  40000
   346    345    259    239     88     67    110     18
 90000 160000 250000  1e+06
     7      4      3      1
```

Hide

```
Break <- sort(unique(Area))[3:11]

d <- numeric(9)
for (i in 1:9) {
  model <-
    lm(log(Species)~(Area<Break[i])*log(Area)+(Area>=Break[i])*log(Area))
  d[i] <- summary(model)[[6]] }

windows(7,4)
```

Hide

```
par(mfrow=c(1,2))
plot(log(Break),d,typ="l",col="red")
```

Hide

```
Break[which(d==min(d))]
```

```
[1] 100
```

Hide

```
model2 <- lm(log(Species)~log(Area)*(Area<100)+log(Area)*(Area>=100))
anova(model1,model2)
```

```
Analysis of Variance Table

Model 1: log(Species) ~ log(Area)
Model 2: log(Species) ~ log(Area) * (Area < 100) + log(Area) * (Area >=
    100)
  Res.Df    RSS Df Sum of Sq      F    Pr(>F)
1   1485 731.98
2   1483 631.36  2    100.62 118.17 < 2.2e-16 ***
---
Signif. codes:
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Hide

```
summary(model2)
```

```
Call:
lm(formula = log(Species) ~ log(Area) * (Area < 100) + log(Area) *
    (Area >= 100))

Residuals:
    Min      1Q  Median      3Q     Max
-2.5058 -0.3091  0.1128  0.4822  1.3443

Coefficients: (2 not defined because of singularities)
                        Estimate Std. Error t value
(Intercept)             1.156087   0.066228  17.456
log(Area)               0.282137   0.009082  31.066
Area < 1001            -0.539270   0.066228  -8.143
Area >= 1001                  NA         NA      NA
log(Area):Area < 1001   0.128057   0.009082  14.100
log(Area):Area >= 1001        NA         NA      NA
                       Pr(>|t|)
(Intercept)             < 2e-16 ***
log(Area)               < 2e-16 ***
Area < 1001            8.12e-16 ***
Area >= 1001                NA
log(Area):Area < 1001   < 2e-16 ***
log(Area):Area >= 1001      NA
---
Signif. codes:
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.6525 on 1483 degrees of freedom
Multiple R-squared:  0.724, Adjusted R-squared:  0.7235
F-statistic:  1297 on 3 and 1483 DF,  p-value: < 2.2e-16
```

Hide

```
summary(model2)[[4]]
```

```
                        Estimate  Std. Error   t value
(Intercept)            1.1560865 0.066227861 17.456196
log(Area)              0.2821369 0.009081866 31.065965
Area < 1001           -0.5392698 0.066227861 -8.142642
log(Area):Area < 1001  0.1280573 0.009081866 14.100333
                           Pr(>|t|)
(Intercept)            3.314281e-62
log(Area)             1.267902e-163
Area < 1001            8.117406e-16
log(Area):Area < 1001  1.834740e-42
```

Hide

```
a1 <- summary(model2)[[4]][1]+summary(model2)[[4]][3]
a2 <- summary(model2)[[4]][1]
b1 <- summary(model2)[[4]][2]+summary(model2)[[4]][4]
b2 <- summary(model2)[[4]][2]

plot(log(Area),log(Species),col="blue")
lines(c(-5,4.6),c(a1+b1*-5,a1+b1*4.6),col="red")
```

Hide

```
lines(c(4.6,15),c(a2+b2*4.6,a2+b2*15),col="red")
```

Section 10.13.1

Hide

```
setwd("C:/Users/Nathan/Desktop/school/statistical data management/therbook")
```

```
The working directory was changed to C:/Users/Nathan/Desktop/school/statistical data management/therbook inside a notebook chunk. The working directory will be reset when the chunk is finished running. Use the knitr root.dir option in the setup chunk to change the working directory for notebook chunks.
```

Hide

```
ozone.pollution <- read.table("ozone.data.txt",header=T)
attach(ozone.pollution)
names(ozone.pollution)
```

```
[1] "rad"   "temp"  "wind"  "ozone"
```

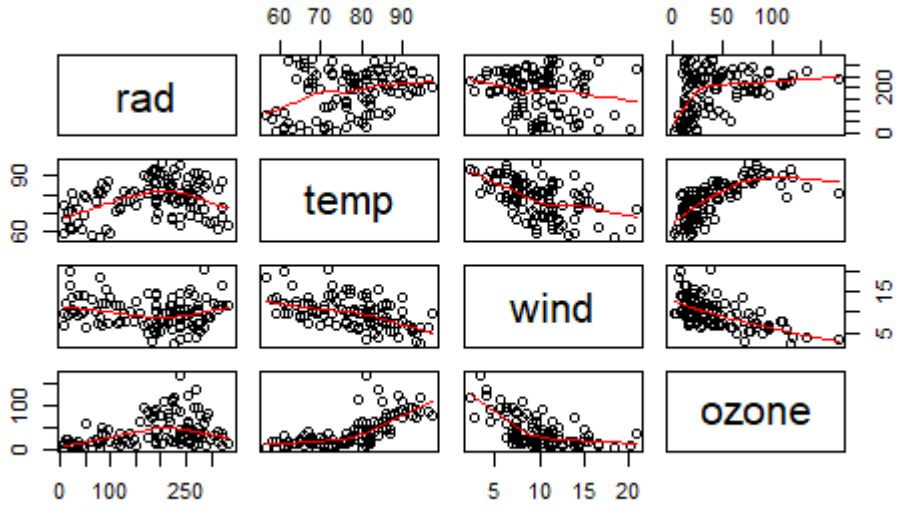
Hide

```
pairs(ozone.pollution,panel=panel.smooth)

library(mgcv)
```

```
Loading required package: nlme
This is mgcv 1.8-28. For overview type 'help("mgcv-package")'.
```

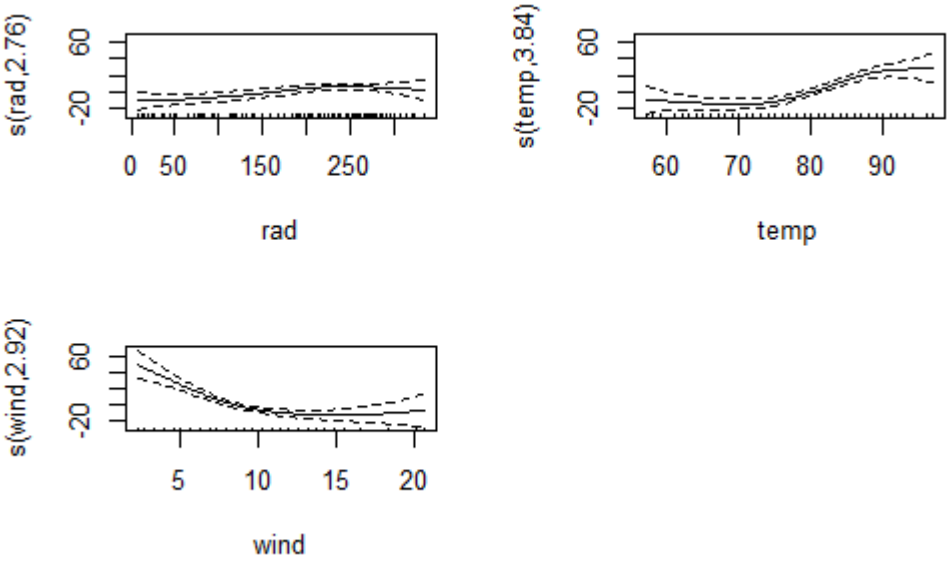
Hide

```
par(mfrow=c(2,2))
```



Hide

```
model <- gam(ozone~s(rad)+s(temp)+s(wind))
plot(model)

library(tree)
model <- tree(ozone~.,data=ozone.pollution)
par(mfrow=c(1,1))
```



Hide

```
plot(model)
text(model)
```

Hide

```
w2 <- wind^2
t2 <- temp^2
r2 <- rad^2
tw <- temp*wind
wr <- wind*rad
tr <- temp*rad
wtr <- wind*temp*rad

model1 <- lm(ozone~rad+temp+wind+t2+w2+r2+wr+tr+tw+wtr)
summary(model1)
```

```
Call:
lm(formula = ozone ~ rad + temp + wind + t2 + w2 + r2 + wr +
    tr + tw + wtr)

Residuals:
    Min      1Q  Median      3Q     Max
-38.894 -11.205  -2.736   8.809  70.551

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  5.683e+02  2.073e+02   2.741  0.00725 **
rad         -3.117e-01  5.585e-01  -0.558  0.57799
temp        -1.076e+01  4.303e+00  -2.501  0.01401 *
wind        -3.237e+01  1.173e+01  -2.760  0.00687 **
t2           5.833e-02  2.396e-02   2.435  0.01668 *
w2           6.106e-01  1.469e-01   4.157 6.81e-05 ***
r2          -3.619e-04  2.573e-04  -1.407  0.16265
wr           2.054e-02  4.892e-02   0.420  0.67552
tr           8.403e-03  7.512e-03   1.119  0.26602
tw           2.377e-01  1.367e-01   1.739  0.08519 .
wtr         -4.324e-04  6.595e-04  -0.656  0.51358
---
Signif. codes:
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 17.82 on 100 degrees of freedom
Multiple R-squared:  0.7394,    Adjusted R-squared:  0.7133
F-statistic: 28.37 on 10 and 100 DF,  p-value: < 2.2e-16
```

Hide

```
model2 <- update(model1,~.-wtr)
summary(model2)
```

```
Call:
lm(formula = ozone ~ rad + temp + wind + t2 + w2 + r2 + wr +
    tr + tw)

Residuals:
    Min      1Q  Median      3Q     Max
-39.611 -11.455  -2.901   8.548  70.325

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  5.245e+02  1.957e+02   2.680   0.0086 **
rad          2.628e-02  2.142e-01   0.123   0.9026
temp        -1.021e+01  4.209e+00  -2.427   0.0170 *
wind        -2.802e+01  9.645e+00  -2.906   0.0045 **
t2           5.953e-02  2.382e-02   2.499   0.0141 *
w2           6.173e-01  1.461e-01   4.225 5.25e-05 ***
r2          -3.388e-04  2.541e-04  -1.333   0.1855
wr          -1.127e-02  6.277e-03  -1.795   0.0756 .
tr           3.750e-03  2.459e-03   1.525   0.1303
tw           1.734e-01  9.497e-02   1.825   0.0709 .
---
Signif. codes:
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 17.77 on 101 degrees of freedom
Multiple R-squared:  0.7383,    Adjusted R-squared:  0.715
F-statistic: 31.66 on 9 and 101 DF,  p-value: < 2.2e-16
```

Hide

```
model3 <- update(model2,~.-r2)
summary(model3)
```

```
Call:
lm(formula = ozone ~ rad + temp + wind + t2 + w2 + wr + tr + 
    tw)

Residuals:
    Min      1Q  Median      3Q     Max 
-39.188 -11.387  -1.500   8.752  71.289 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 486.346603 194.333075   2.503  0.01392 *  
rad          -0.043163   0.208535  -0.207  0.83644    
temp         -9.446780   4.185240  -2.257  0.02613 *  
wind        -26.471461   9.610816  -2.754  0.00697 ** 
t2            0.056966   0.023835   2.390  0.01868 *  
w2            0.599709   0.146069   4.106 8.14e-05 ***
wr           -0.011359   0.006300  -1.803  0.07435 .  
tr            0.003160   0.002428   1.302  0.19600    
tw            0.157637   0.094595   1.666  0.09869 .  
---
Signif. codes:  
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 17.83 on 102 degrees of freedom
Multiple R-squared:  0.7337,    Adjusted R-squared:  0.7128 
F-statistic: 35.12 on 8 and 102 DF,  p-value: < 2.2e-16
```

Hide

```
model4 <- update(model3,~.-tr)
summary(model4)
```

```
Call:
lm(formula = ozone ~ rad + temp + wind + t2 + w2 + wr + tw)

Residuals:
    Min      1Q  Median      3Q     Max
-41.379 -11.375  -2.217   8.921  71.247

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) 514.401470 193.783580   2.655  0.00920 **
rad           0.212945   0.069283   3.074  0.00271 **
temp        -10.654041   4.094889  -2.602  0.01064 *
wind        -27.391965   9.616998  -2.848  0.00531 **
t2            0.067805   0.022408   3.026  0.00313 **
w2            0.619396   0.145773   4.249 4.72e-05 ***
wr           -0.013561   0.006089  -2.227  0.02813 *
tw            0.169674   0.094458   1.796  0.07538 .
---
Signif. codes:
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 17.89 on 103 degrees of freedom
Multiple R-squared:  0.7292,    Adjusted R-squared:  0.7108
F-statistic: 39.63 on 7 and 103 DF,  p-value: < 2.2e-16
```

Hide

```
model5 <- update(model4,~.-tw)
summary(model5)
```

```
Call:
lm(formula = ozone ~ rad + temp + wind + t2 + w2 + wr)

Residuals:
    Min      1Q  Median      3Q     Max
-44.478 -10.735  -2.437   9.685  77.543

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) 223.573855 107.618223   2.077 0.040221 *
rad           0.173431   0.066398   2.612 0.010333 *
temp         -5.197139   2.775039  -1.873 0.063902 .
wind        -10.816032   2.736757  -3.952 0.000141 ***
t2            0.043640   0.018112   2.410 0.017731 *
w2            0.430059   0.101767   4.226 5.12e-05 ***
wr           -0.009819   0.005783  -1.698 0.092507 .
---
Signif. codes:
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 18.08 on 104 degrees of freedom
Multiple R-squared:  0.7208,    Adjusted R-squared:  0.7047
F-statistic: 44.74 on 6 and 104 DF,  p-value: < 2.2e-16
```

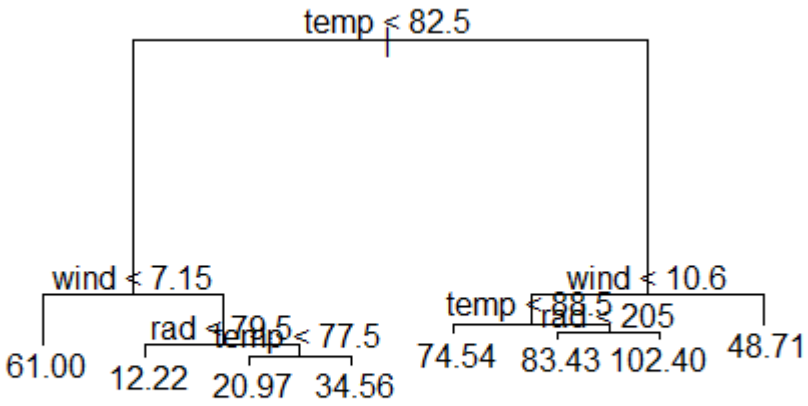
Hide

```
model6 <- update(model5,~.-wr)
summary(model6)
```

```
Call:
lm(formula = ozone ~ rad + temp + wind + t2 + w2)

Residuals:
    Min      1Q  Median      3Q     Max
-48.044 -10.796  -4.138   8.131  80.098

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 291.16758  100.87723   2.886  0.00473 **
rad           0.06586    0.02005   3.285  0.00139 **
temp         -6.33955    2.71627  -2.334  0.02150 *
wind        -13.39674    2.29623  -5.834 6.05e-08 ***
t2            0.05102    0.01774   2.876  0.00488 **
w2            0.46464    0.10060   4.619 1.10e-05 ***
---
Signif. codes:
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 18.25 on 105 degrees of freedom
Multiple R-squared:  0.713, Adjusted R-squared:  0.6994
F-statistic: 52.18 on 5 and 105 DF,  p-value: < 2.2e-16
```

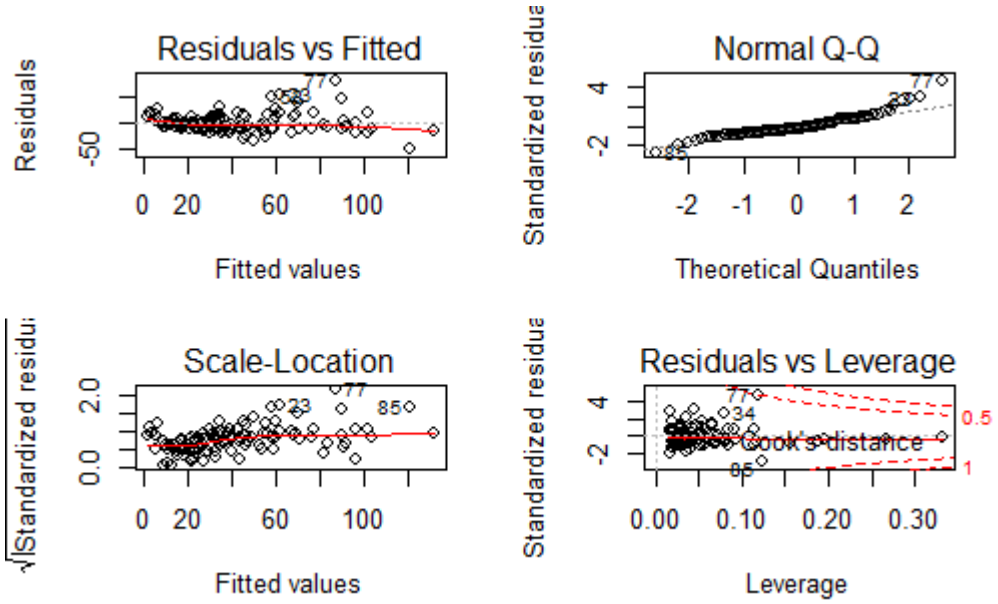
Hide

```
par(mfrow=c(2,2))
```



Hide

```
plot(model6)
```



Hide

```
model7 <- lm(log(ozone) ~ rad+temp+wind+t2+w2+r2+wr+tr+tw+wtr)
summary(model7)
```

```
Call:
lm(formula = log(ozone) ~ rad + temp + wind + t2 + w2 + r2 +
    wr + tr + tw + wtr)

Residuals:
     Min       1Q   Median       3Q      Max
-1.91943 -0.24169 -0.01742  0.28213  1.11802

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.803e+00  5.676e+00   0.494   0.6225
rad          2.771e-02  1.529e-02   1.812   0.0729 .
temp        -3.018e-02  1.178e-01  -0.256   0.7983
wind        -9.812e-02  3.211e-01  -0.306   0.7605
t2           6.034e-04  6.559e-04   0.920   0.3598
w2           8.732e-03  4.021e-03   2.172   0.0322 *
r2          -1.489e-05  7.043e-06  -2.114   0.0370 *
wr          -2.001e-03  1.339e-03  -1.494   0.1382
tr          -2.507e-04  2.056e-04  -1.219   0.2256
tw          -1.985e-03  3.742e-03  -0.530   0.5971
wtr          2.535e-05  1.805e-05   1.404   0.1634
---
Signif. codes:
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4877 on 100 degrees of freedom
Multiple R-squared:  0.7116,    Adjusted R-squared:  0.6827
F-statistic: 24.67 on 10 and 100 DF,  p-value: < 2.2e-16
```

Hide

```
model8 <- update(model7,~.-wtr)
summary(model8)
```

```

Call:
lm(formula = log(ozone) ~ rad + temp + wind + t2 + w2 + r2 +
    wr + tr + tw)

Residuals:
     Min       1Q   Median       3Q      Max
-1.99582 -0.24838 -0.04271  0.32080  1.07835

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  5.373e+00  5.398e+00   0.995   0.3219
rad          7.896e-03  5.908e-03   1.336   0.1844
temp        -6.230e-02  1.161e-01  -0.537   0.5927
wind        -3.531e-01  2.660e-01  -1.327   0.1874
t2           5.332e-04  6.571e-04   0.811   0.4191
w2           8.340e-03  4.030e-03   2.069   0.0411 *
r2          -1.624e-05  7.010e-06  -2.317   0.0225 *
wr          -1.368e-04  1.731e-04  -0.790   0.4313
tr           2.195e-05  6.783e-05   0.324   0.7469
tw           1.784e-03  2.620e-03   0.681   0.4975
---
Signif. codes:
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4901 on 101 degrees of freedom
Multiple R-squared:  0.7059,    Adjusted R-squared:  0.6797
F-statistic: 26.93 on 9 and 101 DF,  p-value: < 2.2e-16
```

Hide

```
model9 <- update(model8,~.-tr)
summary(model9)
```

```
Call:
lm(formula = log(ozone) ~ rad + temp + wind + t2 + w2 + r2 +
    wr + tw)

Residuals:
     Min       1Q   Median       3Q      Max
-1.96263 -0.24298 -0.04081  0.31953  1.09081

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  5.516e+00  5.357e+00   1.030  0.30558
rad          9.533e-03  3.036e-03   3.140  0.00221 **
temp        -6.949e-02  1.134e-01  -0.613  0.54157
wind        -3.574e-01  2.645e-01  -1.351  0.17966
t2           6.029e-04  6.180e-04   0.976  0.33160
w2           8.451e-03  3.998e-03   2.114  0.03697 *
r2          -1.584e-05  6.865e-06  -2.307  0.02310 *
wr          -1.517e-04  1.662e-04  -0.913  0.36341
tw           1.846e-03  2.601e-03   0.710  0.47956
---
Signif. codes:
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4879 on 102 degrees of freedom
Multiple R-squared:  0.7056,    Adjusted R-squared:  0.6825
F-statistic: 30.56 on 8 and 102 DF,  p-value: < 2.2e-16
```

Hide

```
model10 <- update(model9,~.-tw)
summary(model10)
```

```
Call:
lm(formula = log(ozone) ~ rad + temp + wind + t2 + w2 + r2 + 
    wr)

Residuals:
     Min       1Q   Median       3Q      Max 
-1.89186 -0.26391 -0.03075  0.33076  1.09627 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.326e+00  2.907e+00   0.800  0.42544
rad          8.875e-03  2.884e-03   3.077  0.00268 **
temp        -9.290e-03  7.515e-02  -0.124  0.90185
wind        -1.772e-01  7.366e-02  -2.405  0.01795 *
t2           3.360e-04  4.892e-04   0.687  0.49375
w2           6.389e-03  2.739e-03   2.333  0.02162 *
r2          -1.515e-05  6.781e-06  -2.235  0.02761 *
wr          -1.112e-04  1.557e-04  -0.714  0.47676
---
Signif. codes:
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4867 on 103 degrees of freedom
Multiple R-squared:  0.7041,    Adjusted R-squared:  0.684 
F-statistic: 35.02 on 7 and 103 DF,  p-value: < 2.2e-16
```

Hide

```
model11 <- update(model10,~.-t2)
summary(model11)
```

```
Call:
lm(formula = log(ozone) ~ rad + temp + wind + w2 + r2 + wr)

Residuals:
     Min       1Q   Median       3Q      Max
-1.82031 -0.25479 -0.02779  0.33595  1.15024

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.989e-01  7.571e-01   0.527  0.59936
rad          8.996e-03  2.871e-03   3.133  0.00225 **
temp         4.214e-02  6.246e-03   6.746 8.79e-10 ***
wind        -1.816e-01  7.320e-02  -2.481  0.01472 *
w2           6.758e-03  2.679e-03   2.523  0.01316 *
r2          -1.477e-05  6.740e-06  -2.191  0.03071 *
wr          -1.368e-04  1.507e-04  -0.908  0.36615
---
Signif. codes:
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4855 on 104 degrees of freedom
Multiple R-squared:  0.7028,    Adjusted R-squared:  0.6856
F-statistic: 40.98 on 6 and 104 DF,  p-value: < 2.2e-16
```

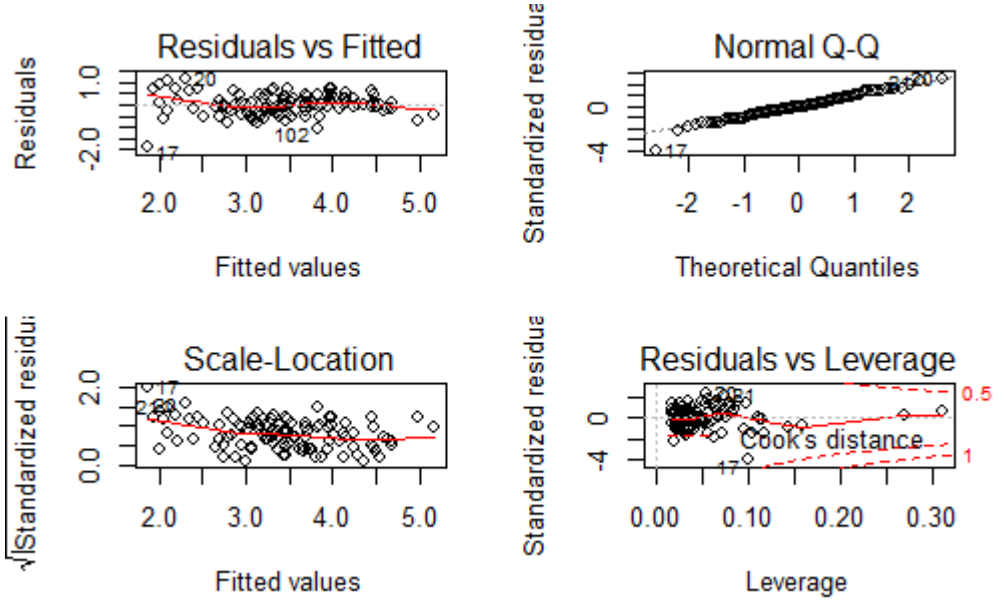
Hide

```
model12 <- update(model11,~.-wr)
summary(model12)
```

```
Call:
lm(formula = log(ozone) ~ rad + temp + wind + w2 + r2)

Residuals:
     Min       1Q   Median       3Q      Max
-1.85551 -0.25578  0.00248  0.31349  1.16251

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  7.724e-01  6.350e-01   1.216 0.226543
rad          7.466e-03  2.323e-03   3.215 0.001736 **
temp         4.193e-02  6.237e-03   6.723 9.52e-10 ***
wind        -2.211e-01  5.874e-02  -3.765 0.000275 ***
w2           7.390e-03  2.585e-03   2.859 0.005126 **
r2          -1.470e-05  6.734e-06  -2.183 0.031246 *
---
Signif. codes:
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4851 on 105 degrees of freedom
Multiple R-squared:  0.7004,    Adjusted R-squared:  0.6861
F-statistic:  49.1 on 5 and 105 DF,  p-value: < 2.2e-16
```

Hide

```
plot(model12)
```



Hide

```
NA
NA
```